

# Gerald Amiel Ballena

✉ gmballena@up.edu.ph     github.com/GABallena     linkedin.com/in/gerald-amiel-ballena  
📍 Metro Manila, Philippines     orcid.org/0009-0000-8857-9755

## Professional Summary

---

Versatile Bioinformatics Engineer with extensive experience in computational biology, bioinformatics pipelines, and data analysis. Highly skilled in automating workflows using Snakemake and Python, optimizing tools for high-throughput sequence analysis, and ensuring reproducibility and efficiency. Adept at collaborating with research teams to produce clear, reliable solutions while solving complex bioinformatics challenges. Proficient in bioinformatics tools and skilled in data visualization and pipeline development.

## Experience

---

2024  Present

Project Technical Specialist

University of the Philippines, College of Public Health

- Developed bioinformatics workflows and automated pipelines for large-scale sequencing data analysis.
- Collaborated with teams across public health, microbiology, and environmental science to produce clear, reproducible research data.
- Wrote Python scripts to automate data processing, error-checking, and visualizations, streamlining data management and reducing errors by 90%.
- Enhanced the analysis pipeline, leading to actionable insights for public health surveillance and research.

## Education

---

University of the Philippines Diliman .....

Graduated: July 2022

Thesis: In silico assessment of the association of pathogenicity and metal-resistance potential of *Fusarium* spp.  
Pre-print Link

Accomplishments: DOST ASTHRDP-Scholarship

Calculated Effective GPA (MS): 1.72

University of the Philippines Baguio .....

Graduated: June 2018

Thesis: Bioelectrocatalysis by Novel Electrogenic Alkaliphilic Bacteria *Bacillus* sp. BAB-3442 Using Dual-Chambered Microbial Fuel Cell ..... Poster Presentation (PSM 47)

## Skills

---

Technical Skills .....

Programming Languages: Python (Fluent), R (Advanced), Bash (Intermediate), Perl & BioPerl (Intermediate), C++ (Fair).

Bioinformatics Tools: Kraken2, MetaPhlAn, MEGAHIT, Snakemake, Prokka, Jellyfish, BUSCO, and others.

Workflow Automation: Snakemake, Conda, YAML (Configuration Files).

Data Visualization: ggplot2, Plotly, matplotlib, QGIS.

High-Performance Computing: Slurm, Docker.

## Soft Skills.....

**Collaboration:** Proficient in working with research teams, ensuring smooth project execution and data delivery.

**Problem-Solving:** Skilled at troubleshooting issues within bioinformatics pipelines and proposing effective solutions.

**Technical Writing:** Expert in preparing clear technical reports and scientific documentation, focusing on precision and clarity.

## Certifications and Relevant Coursework

---

### Certifications.....

**Fundamentals:** AI Fundamentals ..... 2024

**Fundamentals:** Data Literacy ..... 2024

**Advanced:** PH527x: Principles, Statistical and Computational Tools for Reproducible Data Science .... 2024

### Relevant Courses.....

**Introductory:** Introduction to Data Engineering ..... 2020

**Introductory:** COVID-19 Contact Tracing ..... 2020

**Intermediate:** Biostatistics in Public Health ..... 2020

**Intermediate:** Genomic Data Science ..... 2024

**Intermediate:** Genomic Analysis track ..... 2024

**Advanced:** Hyperparameter Tuning in R, Designing Machine Learning Workflows in Python ..... 2024

**Advanced:** Bioinformatics via Coursera ..... 2024

(*University of California, San Diego*)

- Finding Hidden Messages (with Honors)

### Professional Development.....

**Ongoing:** Ensemble Methods in Python (Bagging, Boosting, Stacking) Introduction to AWS Supervised Learning with scikit-learn

## Scripts and Workflows

---

**Key Pipelines and Workflows:** Developed and implemented scalable workflows and pipelines using Snakemake, Python, and Bash for metagenomic analysis, diversity profiling, and bioinformatics tool management. Highlights include:

- **Metagenomic Analysis Pipeline:** Automated workflows for trimming, taxonomic profiling (**Kraken2**, **Bracken**), and diversity calculations (**scikit-bio**), ensuring high reproducibility and scalability.
- **Comprehensive Binning Workflow:** Designed workflows for assembly, binning (**MetaWRAP**, **MEGAHIT**), and MAG validation (**CheckM2**), streamlining large-scale metagenomic projects.
- **Plasmid and ARG Analysis:** Built pipelines for plasmid detection and antimicrobial resistance profiling using **metaSPAdes**, **PlasmidFinder**, and **RGI**.
- **Diversity Analysis:** Developed Python-based scripts and R workflows for alpha/beta diversity metrics (Shannon, Chao1, Bray-Curtis) and visualizations using **ggplot2**.